

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.

MC
5
0400 06-05-01
(TRANSLATION)

Japanese Unexamined Patent Publication No. HEI 2-76072

Publication Date : March 15, 1990

Application No.: SHO 63-227502

Filing Date : September 13, 1988

Applicant : KABUSHIKI-KAISYA RICOH

Inventor : Tetsuya MORITA

Title of the Invention : SEARCH APPARATUS FOR VARIABLE
LONG CHARACTER STRINGS

Remarks: This reference is discussed in the specification of the subject application.

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 02-076072
(43)Date of publication of application : 15.03.1990

(51)Int.Cl. G06F 15/40
G06F 15/40

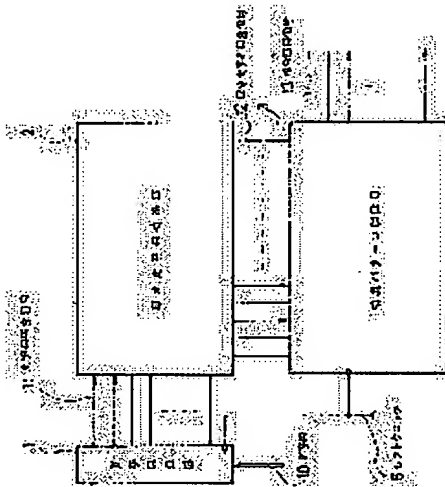
(21)Application number : 63-227502 (71)Applicant : RICOH CO LTD
(22)Date of filing : 13.09.1988 (72)Inventor : MORITA TETSUYA

(54) VARIABLE LENGTH CHARACTER STRING DETECTOR

(57)Abstract:

PURPOSE: To compress the scale of hardware and to detect all patterns of the continuous error, omission, and insertion of a character by providing a character collation part and a partial character string detecting part.

CONSTITUTION: The character collation part 1 registers a detected character string 10 as a registration character string, and outputs a character collation signal 11 at every character representing whether or not each character of the character string inputted from a regular document file is present in the registration character string. The partial character string detecting part 2 outputs a character string collation signal 12 representing the presence/absence of the registration character string and its continuous partial character string by taking the AND of the character collation signal 11. A detecting pattern control part 3 outputs a collation result signal 13 representing an error character string including a continuous error character by taking the AND of the collation result signal 13 representing the registration character string, the not of the signal, and another partial character string signal in the partial character string collation signal 12. In such a way, the error character string including the registration character string and the continuous error character can be detected.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

⑫ 公開特許公報(A) 平2-76072

⑤ Int.Cl.³

G 06 F 15/40

識別記号

5 0 0 E
5 1 0 L

庁内整理番号

7313-5B
7313-5B

④ 公開 平成2年(1990)3月15日

審査請求 未請求 請求項の数 2 (全8頁)

⑥ 発明の名称 可変長文字列検出装置

② 特 願 昭63-227502

② 出 願 昭63(1988)9月13日

⑦ 発 明 者 森 田 哲 也 東京都大田区中馬込1丁目3番6号 株式会社リコー内
 ⑦ 出 願 人 株 式 会 社 リ コ ー 東京都大田区中馬込1丁目3番6号
 ⑦ 代 理 人 弁 理 士 香 取 孝 雄 外1名

明 細 書

1. 発明の名称

可変長文字列検出装置

2. 特許請求の範囲

1. それぞれ対応する複数の文字レジスタを備え、検出する文字列が予め登録文字列として登録される登録レジスタ、およびファイルから入力される文字列をクロック信号に同期して一文字づつ前記文字レジスタに移送するシフトレジスタを備え、対応する前記文字レジスタの内容を比較し、前記文字列と前記登録文字列の一致の有無を示す1文字ごとの文字列照合信号を出力する文字照合部と、

前記文字列照合信号に所定の論理演算処理を施し、その結果より前記登録文字列およびその連続する部分文字列の有無を示す部分文字列照合信号を出力する部分文字列検出部と、

前記部分文字列照合信号に所定の論理演算処理を施し、その結果を照合結果信号として出力する検出パターン制御部とを有し、

前記照合結果信号により、前記登録文字列の有無および前記部分文字列と連続する誤り文字列よりなる各種の誤り文字列を前記ファイルから検出することを特徴とする可変長文字列検出装置。

2. 請求項1に記載の装置において、前記検出パターン制御部は、前記クロック信号に同期して順番に導通状態となる少なくとも1つの遅延素子を含み、該素子を通過した前記部分文字列照合信号に所定の論理演算処理を施した結果を照合信号として出力し、該照合信号により、前記登録文字列から連続する文字欠落、または連続する余剰文字を含む誤り文字列を前記ファイルから検出することを特徴とする可変長文字列検出装置。

3. 発明の詳細な説明

(産業上の利用分野)

本発明は、磁気ディスクや光ディスク等の記憶媒体に保存されている文書ファイルを入力し、あらかじめ登録されている文字列と照合を行

うことにより、文字情報(テキスト)中から登録文字列および誤り文字列を検索する装置に関する。

(従来の技術)

従来の文字列検出方法としては、(1)ソートサーチ法、(2)連想メモリ法、(3)セルラアレイ法、(4)有限状態オートマトン法、(5)DP法等があげられるが、これらの方法には以下のような問題があった。(2)、(3)は、長い文字列が扱えない、(1)、(2)、(3)、(5)は可変長文字列のノンアンカーマッチが扱えない、(5)は処理速度が遅い、(3)、(4)、(5)はハードウェアサイズが小さくならないなどの問題がある。これらの問題点の一部を解決するために、(2)と(4)を組み合わせた文字列検索LSI(日本電気-ISSP)が開発されている。これを示す文献に、高橋、他「ストリング・マッチング・ハードウェアのアーキテクチャ」電子通信学会研究報告(コンピュータシステム)、CPSY 86-57(1986.7)がある。この場合でも、装置規模が許容される誤り文字数に比例す

るため、ハードウェア構成の制約により現実には誤り文字数が1文字程度のLSIしか製作されていない。

(発明が解決しようとする課題)

上述の文献のISSPはN文字の誤り・抜け・挿入のすべてのパターンを検出するように構成できるが、現実的には伝送路エラーや2次記憶で発生するエラーはバースト的な誤りがほとんどであり、ランダム誤りは少ない。そこで本発明では、連続するN文字の誤り・抜け・挿入のエラーのみを検出し、連続するエラーのみを扱うことによりハードウェアの規模を小さくし、実用的な可変長文字列検出装置を提供することを目的とする。

(課題を解決するための手段)

本発明による可変長文字列検出装置は、それぞれ対応する複数の文字レジスタを備え、検出する文字列が予め登録文字列として登録された登録レジスタ、およびファイルから入力された文字列をクロック信号に同期して一文字づつ

字レジスタに移送するシフトレジスタを備え、対応する文字レジスタの内容を比較し、前記文字列と前記登録文字列の一致の有無を示す1文字ごとの文字列照合信号を出力する文字照合部と、文字照合信号に所定の論理演算処理を施し、その結果より登録文字列およびその連続する部分文字列の有無を示す部分文字列照合信号を出力する部分文字列検出部と、部分文字列照合信号に所定の論理演算処理を施し、その結果を照合結果信号として出力する検出パターン制御部とを有し、照合結果信号により、登録文字列の有無および部分文字列と連続する誤り文字列よりなる各種の誤り文字列を前記ファイルから検出することを特徴とする。

(作用)

文字照合部は予め検出する文字列を登録文字列として登録し、文書ファイル等から入力される文字列の各文字が登録文字列に有るか否かを示す文字ごとの文字照合信号を出力する。部分文字列検出部は文字照合信号の論理積をとるこ

とにより、登録文字列とその連続する部分文字列の有無を示す部分文字列照合信号を出力する。検出パターン制御部は部分文字列照合信号の中で、登録文字列を示す照合結果信号と、同信号の否定と他の部分文字列信号の論理積をとることにより、連続する誤り文字を含む誤り文字列を示す照合結果信号を出力する。このようにして、可変長文字列検出装置は登録文字列および連続する誤り文字を含む誤り文字列を検出することができる。

(実施例)

次に、本発明の実施例を添付図面を用いて説明する。第1図は本発明の可変長文字列検出装置の一実施例を示すブロック図である。この装置は、磁気ディスクや光ディスク等の2次記憶に保存されている文書ファイルを入力し、あらかじめ登録されている文字列と照合を行ない登録文字列と同一の文字列の検出、1文字以上の連続する誤り・文字抜け・余剰文字挿入を出力する。

装置は、各部1, 2および3よりなり、各部の入出力動作はシフトクロック5と同期して行われる。文字照合部1は、複数の文字レジスタよりなり、文書ファイル等から検出する文字列が登録文字列として予め登録される。文書ファイル等からの文字列を検出するときは、入力される文字列10と登録文字列の対応する各文字を比較し、一致の有無を示す文字照合信号11を出力する。部分文字列検出部2は、この信号11を入力し、所定の論理演算処理を施し、登録文字列およびその部分文字列の有無を示す部分文字列照合信号12を出力する。検出パターン制御部3は入力される信号12に所定の論理演算処理を施し登録文字列の有無および部分文字列を含む各種の文字列を示す照合結果信号13を出力する。

第2図は文字照合部1の構成例を示す。データセクタ20は、登録時に登録文字列1が入力されると、出力aより登録レジスタ31に出力し、照合時に、文書ファイル等の文字列1が

入力されると、出力bより照合レジスタ32に出力する。クロックセクタ21は入力するシフトクロック5を登録時に登録レジスタ31に照合時に照合レジスタ32に出力する。以上のデータセクタ20およびクロックセクタ21の出力切換は登録/照合セレクト信号6により制御される。

登録レジスタ31および照合レジスタ32は図示のようにそれぞれ複数個直列に接続され、シフトレジスタ31aおよび31bを構成する前者は、登録時、入力される登録文字列1の各文字をシフトクロック5に同期して各登録レジスタ31を通して移送し、登録文字列を登録する。後者は、同様に文書ファイル等からの文字列1を各照合レジスタ32を通して移送する。比較器30は各レジスタ31および32の各出力310および320を入力され、一致したとき文字照合信号11をアクティブにする。なお登録レジスタ31はシフトレジスタ31aに限定されるものではなく、予め登録文字列1を記

憶するROM等であっても差支えない。

第6図は、文字照合部1に登録される文字列の例と本装置によって検出可能な文字列パターンを示したものである。[*]は登録文字列に対して正しくない任意の文字を要す。

部分文字列検出部2の構成例を第3図(a)に示す。同部2は抵抗Rを介して電圧 V_{cc} が印加される縦線(列)と、シフトクロック5に同期した文字列照合信号11が印加される横線(行)と、格子点で行間を接続する同図(b)のオープンコレクタ型バッファ40で構成されている。バッファ40は、文字列照合信号11の論理積をとり、登録文字列およびそのN文字が連続する部分文字列を検出する部分文字列検出信号12を列に出力する。

例えば、登録文字列が"STRING"であるとき検出される部分文字列は"S", "ST", "STR", "STRI", "STIRN", "STRING", "TRING", "RING", "ING", "NG", "G"等がある。

上記の部分文字列が文書ファイル等に存在すると、部分文字列照合信号12がアクティブとなる。

検出パターン制御部3の構成例を第4図(a)に示す。同部3は部分文字列検出部2から部分文字列照合信号12が入力される列と、抵抗Rを介して電圧 V_{cc} が印加される行と、格子点に接続される同図(b)のオープンコレクタ型インバートバッファ50およびバッファ40で構成される。バッファ40および50は、列から入力される部分文字列照合信号12のアクティブおよび否定の、論理積をとり照合結果信号13として行に出力する。照合結果信号13がアクティブとなると、信号13の"STRING"は登録文字列、"*TRING"~"STRIN*"は1文字誤りパターンの、また"**RING", "S**ING"は2文字誤りパターン、の文字列が、ファイル文書等にあることを示す。

第5図は検出パターン制御部3の他の構成例であり、登録文字列からN(自然数)字抜け、

またはN字の余剰文字が挿入されたパターンを検出する。直列に接続された1ビットラッチ61および62は、シフトクロック5に同期し、入力Dから入力した部分文字列照合信号12を1クロック遅らせて出力Qに出力して、信号12を1クロック遅延させる遅延素子である。直列に接続されるラッチ61および62の出力Dに接続する列線211および212には、それぞれ1クロックおよび2クロック遅延した部分文字列信号12が出力され、列線211とバッファ40を介して接続される行線に"SRING"～"STRNG"等の一文字抜けパターンが出力され、さらに"STRING"の列線とインバートバッファ50を介して接続する行線に"S*TRING"～"STR*ING"等の一文字挿入パターンが出力される。同様に列線212と接続する行線に"SING"～"STG"等の2文字抜けおよび"S**TRING"～"STR**ING"等の2文字挿入パターンが出力される。このようにして、1文字以上の連続する誤りを検出することができる。

したい文字の文字照合信号11と部分文字列照合信号12をオープンコレクタ型バッファ60を用いて接続することにより部分文字列照合信号12を構成することができる。例えば、"STR"という3文字の文字列を検出したい場合、"S"、"T"、"R"に対応する文字照合信号11と部分文字列照合信号12を接続すればよい。ただしここでいう「接続」とは電氣的等電位にするよう物理的接続を行なうのではなく、論理的にワイヤード・アンドを構成できるようにオープンコレクタバッファ等で接続することである。

第4図(a)の検出パターン制御部3では、同部3に入力された部分文字列検出信号12のうち、必要な信号だけを照合結果信号13と接続することにより図のようにN文字誤りパターンの検出が可能となる。例えば、1文字誤りパターン"S*RING"は"S"に対する部分文字列照合信号12と、"RING"に対する部分文字列照合信号12と、"STRING"に対する部分文字列照合信号12の否定信号を接続することによって構

成に、文書情報から検出すべき文字列(登録文字列)が、"STRING"という文字列である場合を例にして以下説明する。第2図の文字照合部1において、登録/照合セレクト信号6によってデータセレクト20およびクロックセレクト21の出力を登録レジスタ31側に接続し、登録文字列1をシフトクロック5に同期して入力することによって登録が行われる。

可変長文字列検出を行なう際には登録/照合セレクト信号6によってデータセレクト21およびクロックセレクト21の出力を照合レジスタ32側に接続し、文字情報を文書ファイルからシフトクロック5に同期して入力する。入力される文字情報は1文字ごとに比較器30によって登録文字列と比較され、その結果が文字照合信号11として出力される。

第3図の部分文字列検出部2では、行に入力される所定の文字照合信号11の論理積をと列に出力することによって登録文字列1の部分文字列を検出する。第3図に示すように、検

成できる。"STRING"の否定信号を接続するは、本構成例では"S*RING"を満たすパターンのうち正解でない文字列を検出するようにしたためであり、これを接続しなければ正解文字を含むパターン検出が可能となる。

また、第5図に示す構成例における検出パターン制御部3のように部分文字列照合信号1にN段の1ビットラッチ61、62群を付加することによって、N文字抜けパターンおよび文字挿入パターンの検出ができる。例えば、文字抜けパターン"SRING"を検出する信号を得るには1クロック前の"RING"信号と現在"S"信号とを照合結果信号13に接続すればよい。つまり、文字照合部1に入力された文字列が、"SRING"である場合、"RING"が検出された1クロック後に"S"が検出されるためである。2文字以上の抜けパターンについて同様に検出が可能である。

さらに例えば1文字挿入パターン"S*TRING"を検出する信号を得るには、1クロック前の

"S" 信号と現在の "TRING" 信号と現在の "STRING" の否定信号とを照合結果信号 3 に接続すればよい。これも文字照合部 1 に入力された文字列が "S*TRING" である場合、"S" が検出された 1 クロック後に "TRING" が検出されるためである。2 文字以上の挿入パターンの検出についても同様にして検出が可能である。また前述の誤りパターンの例と同様に登録文字列（正解文字列）の信号 "STRING" の否定信号を付加しているが、これを除けば正解文字列を含む検出が可能である。N 文字抜けパターン検出の場合は正解文字列が現われる可能性がないので、正解文字列の否定信号を付加する必要がない。

以上のようにして、登録文字列に対して正解文字列パターン、連続する N 文字誤りパターン、同じく N 文字抜けパターン、同じく N 文字挿入パターンの検出ができる。

（発明の効果）

以上説明したように、本発明によれば、連続

する N 文字の誤り・抜け・挿入のすべてのパターンを検出でき、すべての N 文字の誤りの抜け挿入パターンを検出する装置に比べ、ハードウェア規模が小さくできる。また、同一モジュールの繰返し構造が多いため VLSI 化に好適であるなどの効果がある。

4. 図面の簡単な説明

第 1 図は本発明の可変長文字列検出装置の実施例を示す構成図、

第 2 図は第 1 図に示す実施例における文字照合部のブロック図、

第 3 図(a)は同実施例における部分文字列検出部の構成例を示す図、第 3 図(b)はオープンコレクタ型バッファの記号を説明する図、

第 4 図(a)は同実施例における検出パターン制御部の構成例 1 を示す図、第 4 図(b)はオープンコレクタ型インバートバッファの記号を説明する図、

第 5 図は同実施例における検出パターン制御部の他の構成例を示す図、

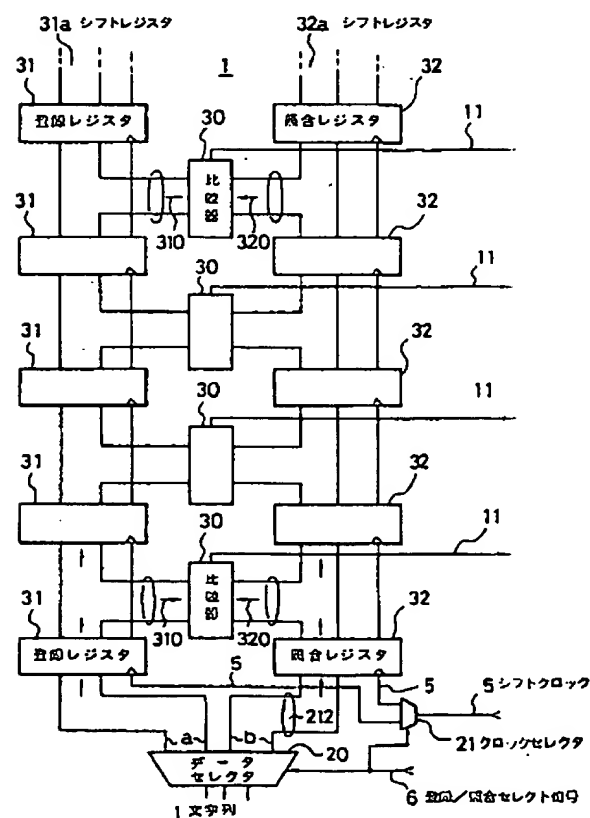
第 6 図は検出可能な文字列のパターンの例を示す図である。

主要部分の符号の説明

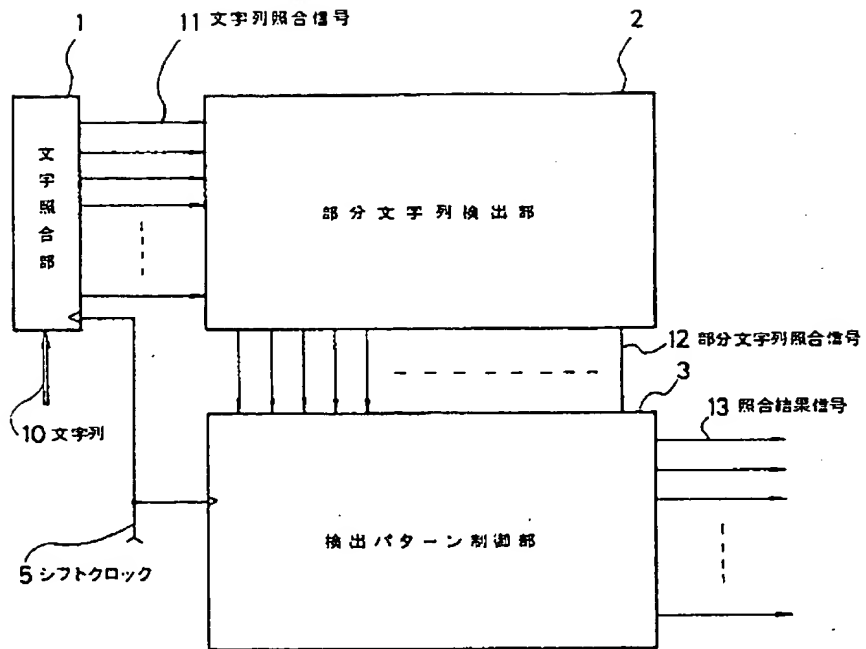
- 1 … 文字照合部
- 2 … 部分文字列検出部
- 3 … 検出パターン制御部
- 20 … データセレクト
- 21 … クロックセレクト
- 30 … 比較器
- 31 … 登録レジスタ
- 32 … 照合レジスタ
- 40 … オープンコレクタ型バッファ
- 50 … オープンコレクタ型インバートバッファ
- 61, 62 … 1 ビットラッチ

特許出願人 株式会社 リコー
代理人 香 取 孝 雄
丸 山 隆 夫

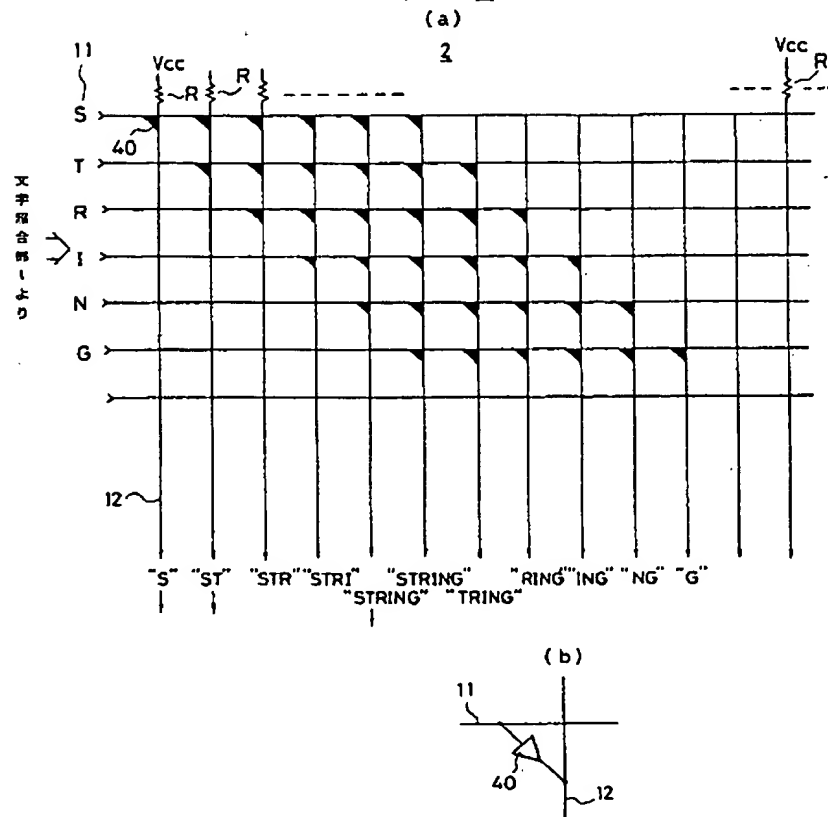
第 2 図

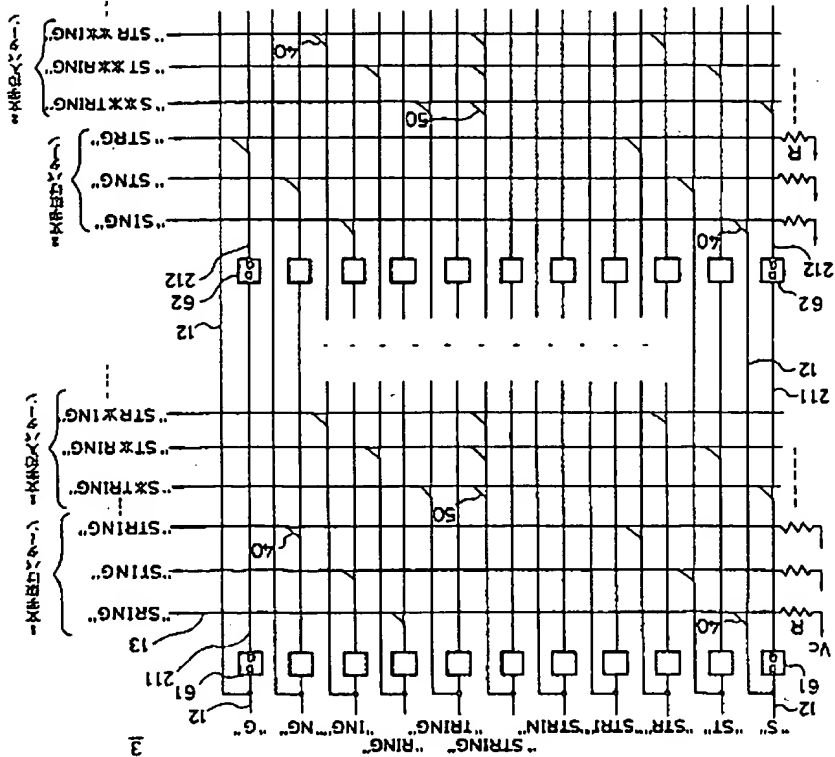


第 1 図

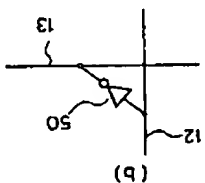


第 3 図

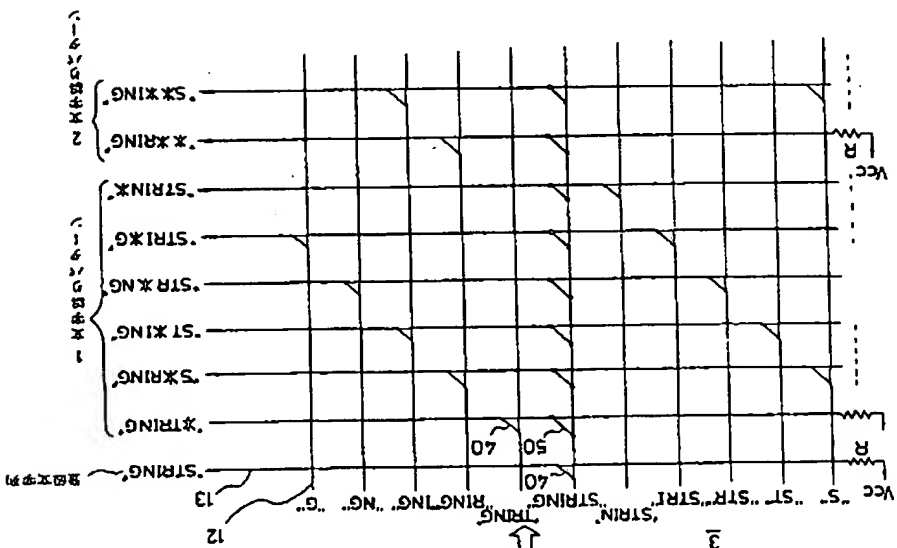




第 5 圖



(b)



第 6 圖 (a)

第 7 圖

第 6 図

置換文字列:

"STRING"

N文字繰り返しパターン:

"XSTRING"

"XXRING"

"SXRING"

"SXXRING"

"STXING"

"STXXING"

"STRXNG"

"STRXXG"

"STRIXG"

"STRIXX"

"STRINX"

(1文字繰り返し)

(2文字繰り返し)

N文字抜けパターン:

"TRING"

"RING"

"SRING"

"SING"

"STING"

"STNG"

"STRNG"

"STRG"

"STRIG"

"STRI"

"STRIN"

(1文字抜け)

(2文字抜け)

N文字挿入パターン:

"SXSTRING"

"SXXSTRING"

"STXRING"

"STXXRING"

"STRXING"

"STRXXING"

"STRIXNG"

"STRIXXG"

"STRINXG"

(1文字挿入)

(2文字挿入)